

ロジスティック回帰分析 (logistic regression)

2008/5/14

経済学研究科

D1 羅嬉穎(ナ・ヒギョン)

ロジスティック回帰分析

1960年代に生活習慣病の原因調査結果を分析する手法として作られた。この手法は医学や薬学、生物学などの自然科学だけでなく社会科学の分野でも利用されるようになった。

(丹後・山岡・高木、1996)

1. ロジステック回帰分析とは

・従属変数が2値変数の際、用いる回帰分析

例：満足・不満足

成功・失敗

発症・非発症

賛成・反対

賛成・どちらでもない・反対の場合は、「賛成する・賛成しない」の2値に落として分析

ある事象の生起の有無

(0と1のデータと考えることが多い)

例：説明変数 従属変数(脳卒中の有無0,1)

性別

飲酒量

GGT

喫煙

血圧etc.

結果では脳卒中になる確率の予測値が示される。(0~1までの確率値)

なぜ回帰分析ではだめなのか

- ・離散変数である2値変数を連続変数として予測するのは無理がある。

(データの分布において正規分布の仮定が不可能)

- ・ x に値を代入した場合、予測値 y^{\wedge} は0 or 1とは限らない。
 $y^{\wedge} = 0.5, -0.6$ etc.(解釈不可能)

確率で考えると

独立変数は従属変数の1という事象が起こる確率に影響するというふうに考えると、

$$P(Y=1) = a + bx$$

確率は $[0, 1]$ の範囲しかとらない。

$y^{\wedge} = 1.6, -0.9$ などは解析不可能

確率のロジットをとると

ロジット(オッズの対数をとったもの)

$$\text{logit}[P(Y = 1)] = \log\left[\frac{p(Y = 1)}{1 - p(Y = 1)}\right]$$

こうすると従属変数の範囲が $[-\infty, \infty]$ になる

ロジスティック回帰分析は生起確率のロジットに
回帰モデルを想定したもの

$$\log\frac{p(Y = 1)}{1 - p(Y = 1)} = a + bx$$

オッズ(odds)

$$\log \frac{p(\mathbf{x})}{1-p(\mathbf{x})} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_r x_r$$

ここから，式変形を行うと

$$\frac{p(\mathbf{x})}{1-p(\mathbf{x})} = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_r x_r)$$

上記左辺を $p(\mathbf{x})$ の $1-p(\mathbf{x})$ に対するオッズといふ

オッズ比にする理由

	A薬	B薬	比(Bの確率/Aの確率)
治癒	90	99	$0.99/0.90=1.1$
未治癒	10	1	$0.01/0.10=0.1$

	A薬	B薬	
治癒	50	55	$0.55/0.50=1.1$
未治癒	50	45	$0.45/0.50=0.9$

「90から99」と「50から55」は評価が異なるべきである。となると、未治癒率で見ればいいということになるが、未治療率で見ると、治癒率と未治癒率の数値が逆になっている場合問題になる。

オッズ比で考えると

	A薬	B薬	オッズ比
治癒	90	99	$\frac{0.99}{1-0.99}$
未治癒	10	1	$\frac{0.90}{1-0.90} = \frac{99}{9} = 11$

	A薬	B薬	オッズ比
治癒	50	55	$\frac{0.55}{1-0.55} = \frac{1.22}{1}$
未治癒	50	45	$\frac{0.50}{1-0.50} = 1$

オッズ比とロジスティック回帰係数 の関係

配布資料参考

オッズ比 = $\text{Exp}(\quad)$

= 説明変数 x が1だけ変化したときの対数
オッズ比

説明変数が1だけ変化するとき、従属変数は
 $\text{Exp}(\quad)$ 倍変化する。

パラメータの推定・検定

ロジスティック回帰分析はロジットの回帰誤差が特殊な分布になる。このため回帰誤差が近似的に正規分布すると見なして、重回帰分析と同じように最小2乗法を利用して回帰分析を行う方法と、最尤法を利用した繰り返し近似計算によって回帰分析を行う方法の、2種類の計算法がある。

SPSSのロジスティック回帰分析でのパラメータの推定・検定には最尤法が用いられる。

最尤値・・・尤度が最も大きくなる値

最尤推定値は正規分布に従う

パラメータの推定・検定

・ワルド検定

推定値を標準誤差で割ったものの絶対値が1.96より大きければ5%水準で有意。

また、 z を2乗したものは自由度1のカイ2乗分布に従う。2乗したものが3.84より大きければ有意。

SPSSでの使い方

1. 「分析」 「回帰」 「二項ロジスティック」を選択
 択
2. 従属変数 従属変数
 説明変数 共変量

参考: 説明変数にカテゴリカル変数があれば、「カテゴリ」を選択し、カテゴリ共変量に移動させてから参照カテゴリを「最初」にして変更を選択する。

引用・参考文献

丹後俊郎・山岡和枝・高木晴良（1996）統計ライブラリー ロジスティック回帰分析 - SASを利用した統計解析の実際 - 朝倉書店

石村貞夫・謝承泰・久保田基夫（2001）よく分かる医学・歯学・薬学のための統計解析 東京図書

鳥居稔 ロジスティック回帰分析

<http://koko15.hus.osaka-.ac.jp/~torii/logistic/ppframe.htm>